K-Fold Cross-Validation:

The procedure has a single parameter called k that refers to the number of groups that a given data sample is to be split into. As such, the procedure is called k-fold cross validation.

This is, to use a limited sample in order to estimate how the model is expected to perform in general when used to make predictions on data not used during the training of the model.

Advantages: (i) simple to understand

(ii) Generally results in a less-biased or less optimistic estimate of the model skill than other methods, such as a simple train/test split.

The general procedure is as follows:

(i) Shuffle the dataset randomly.

(ii) Split the dataset into k groups.

(iii) For each unique group:

(a): Take the group as a hold out or test data set

(b): Take the remaining groups as a training data set

(c): Fit a model on the training set and evaluate it on the test set

(d): Retain the evaluation score and discard the model.

(iv) Summarize the skill of the model using the sample of model evaluation scores.